

Does the Teacher Beat the Test? The Value of the Teacher's Assessment in Predicting Student Ability

Eva Feron¹ · Trudie Schils¹  · Bas ter Weel^{1,2}

Published online: 1 July 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract This research investigates to what extent the subjective teacher's assessment of children's ability predicts children's outcomes in the transition from primary to secondary school in terms of initial track allocation, track switching in the first three years of secondary education and subsequent test scores. We apply micro-data from the Netherlands about cognitive test scores and teacher's assessment in primary schools and about track placement, track switching and test scores in secondary schools. Our estimates suggest that the subjective teacher's assessment is about twice as important as the primary school cognitive test scores for initial track placement in secondary

The programs and data used in this paper are available from the authors upon request. We would like to thank Lex Borghans, Flavio Cunha, Bart Golsteyn, Erzo Luttmer, Marc van der Steeg, Dinand Webbink and Karen van der Wiel for comments on an earlier draft. We are also grateful for useful feedback from an anonymous referee, seminar participants at the CPB, Maastricht University and the University of Philadelphia, conference participants at the *Nederlandse Economen Dag* in Amsterdam, de *Onderwijsresearch Dagen* in Brussels, the ESPE conference in Braga, the IWAE conference in Catanzaro, the EALE in Ljubljana and the EALE-SOLE in Montreal. We also thank the feedback of teachers in Communities of Learning where we discussed our work with the field.

Electronic supplementary material The online version of this article (doi:[10.1007/s10645-016-9278-z](https://doi.org/10.1007/s10645-016-9278-z)) contains supplementary material, which is available to authorized users.

✉ Trudie Schils
t.schils@maastrichtuniversity.nl

Eva Feron
e.feron@maastrichtuniversity.nl

Bas ter Weel
b.ter.weel@cpb.nl

¹ Department of Economics, Maastricht University, Maastricht, The Netherlands

² CPB Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands

school. In addition, the teacher's assessment is more predictive of track allocation in 9th grade compared to cognitive test scores. Next, children who switch tracks are more likely to be placed in tracks based on test scores. Also, test scores in 9th grade are predicted by the subjective teacher's assessment, not by test scores in 6th grade.

Keywords Education · Assessment measures · Tracking

JEL Classification I21

1 Introduction

The weight of cognitive test scores in allocating children to different levels of education has increased over time (e.g., [Kautz et al. 2014](#)). The reason why cognitive test scores are widely used as an assessment measure to sort children is that they are assumed to give objective measures of ability levels. Teacher assessments represent more subjective measures of ability and have been criticized for being biased; for example towards gender, children from disadvantaged families or ethnic minorities (e.g., [Dee 2005](#); [Burgess and Greaves 2013](#); [Fairlie et al. 2014](#)). However, teacher assessments could also be valuable complements to cognitive test scores. Teachers work with children on a daily basis, which allows them to also assess other determinants of ability, such as motivation and classroom behaviour (e.g., [Segal 2008](#)).

This research investigates to what extent a teacher's assessment contains additional information that is useful in determining primary school children's ability level on top of the information provided by cognitive test scores. We make use of a database that contains information about cognitive test scores and teacher assessment in primary school and about initial track placement and subsequent careers in secondary school of Dutch children. The Netherlands has an educational system that involves early tracking (i.e., tracking at the age of 12) after the completion of primary school. Most children start primary school at the age of 4, enter 1st grade at the age of 6 and finish primary school at the age of 12. As of secondary school children are allocated to tracks. The track allocation decision is made by the secondary school. It is based on test scores and teacher assessment. Our empirical analysis benefits from this system because we observe both high-stakes cognitive test scores and teacher assessments at the end of 6th grade and the transition from primary to secondary school during which children are allocated to different (hierarchical) education tracks.

The research strategy involves three steps. The first step is to document whether or not there are non-random differences between cognitive test scores and teacher assessment at the individual level. In the second step we investigate whether or not cognitive test scores or the teacher's assessment is more predictive of track placement in 7th grade and track allocation in 9th grade. In a complementary analysis we document the determinants of math and language test scores in 9th grade by correlating them with both ability signals. In the third step we investigate whether track switchers (between 7th and 9th grade) have been allocated according to the teacher's assessment

or cognitive test scores.¹ The application of the assessment measures by secondary schools to allocate children to different tracks (or the lack thereof) yields information about the usefulness of objective and subjective assessment measures in allocating children.

Our database for the empirical analysis includes 4500 children. It includes administrative data from school tracking systems and survey data of children in 6th and 9th grade in the period 2009–2012 (i.e., these children were in 6th grade in 2009 and were followed until 9th grade in 2012). More specifically, the data include both objective and subjective assessment measures in 6th grade, track placement from 7th to 9th grade, test scores on an identical test across all different tracks in 9th grade and measures about demographic factors as well as socio-economic status. The objective assessment measure is a high-stakes standardized test score (so-called *Cito Eindtoets*) which children take in 6th grade. This objective test score provides a well-defined measure of levels of achievement and is annually taken by almost all 6th grade children in the Netherlands.² The subjective assessment measure is the teacher's assessment of the child's ability level. This subjective assessment is made by teachers in 6th grade after they have observed the test score from the objective assessment.

Our four most important findings can be summarized as follows. First, for 19 % of our sample we observe a substantial difference between the objective and subjective assessment measure of ability in 6th grade. In three quarters of these cases the teacher's assessment is higher than the test score. We find that there are systematic differences between the objective and subjective assessment measure. Our most important findings relate to gender and social-economic status. Girls are more likely to receive a teacher assessment that is higher than their test score compared to boys and children from families with lower socio-economic status are less likely to receive a teacher assessment that is higher than the test score compared to children from higher level families.³

Second, our estimates suggest that the teacher's assessment is twice as powerful to explain the gap between the lowest and the highest track placement compared to the test score, in both 7th and 9th grade. These results suggest that secondary schools put more emphasis on the subjective assessment measure relative to the objective assessment measure when allocating children to tracks in 7th grade. It also suggests that the teacher's assessment is not only more predictive of initial track placement but also of the longer term career in secondary school compared to the cognitive test

¹ The benefits of an early tracking system are highest when children are allocated to tracks most suitable given their ability level and that they stay within this track throughout their secondary school careers. These seem to be the most important reasons for early tracking in for example Germany (e.g., [Woessmann 2004](#)) and the Netherlands (e.g., [Diris 2012](#)), two countries that track children the earliest among all OECD countries.

² Approximately 85 % of all children who are in 6th grade participate in the *Cito Eindtoets* every year.

³ These numbers are consistent with previous studies using data about the Dutch education system. For example, [Timmermans et al. \(2012\)](#) investigate the relationship between assessment measures and secondary school careers. They use a different dataset, which contains information about children from all regions in the Netherlands but no information on track placement and switching. Their descriptive statistics about assessment measures are similar to ours, which suggests that our analysis is likely to be representative for the Netherlands. [Golsteyn and Schils \(2014\)](#) show that boys and girls score differently on the 6th grade tests according to differences in personality that might also be related to higher teacher assessments.

score. In addition, secondary schools seem to allocate children in accordance with the highest assessment signal of ability. Finally, our estimates are conservative or lower-bound estimates as we only consider eight different educational tracks in secondary education for these analyses and are very strict in labelling signals as different.⁴

Third, we observe that the teacher's assessment has non-random deviations from the test score. The question is whether these non-random deviations are efficient with respect to later outcomes, such as switching tracks. We observe that approximately 24 % of our sample makes a switch between tracks between 7th and 9th grade. Our analysis suggests that children who are allocated according to the teacher's assessment are the least likely to switch tracks.

Finally, the estimates suggest that the teacher's assessment positively correlates with the 9th grade test scores, whereas the cognitive test score in 6th grade does not explain test scores in 9th grade (when controlling for the teacher's assessment). Switching tracks between 7th and 9th grade seems to have a negative effect on the test score in 9th grade, pointing towards costs of switching.

This paper contributes to the literature about the consequences of using objective and subjective assessment measures for tracking and successive performance. [Dee \(2005\)](#), [Lindahl \(2007\)](#), [Lavy \(2008\)](#), [Gibbons and Chevalier \(2008\)](#), [Cornwell et al. \(2012\)](#) and [Burgess and Greaves \(2013\)](#) all use objective and subjective assessment measures to study discrimination and uncertainty. It is shown that systematic differences exist between these two types of instruments in the assessment of children's performance, such as between boys and girls, or between blacks and whites. [Bernardi et al. \(2014\)](#) show that additional information from cognitive and non-cognitive tests is able to help children make a more efficient track allocation choice. Our contribution to this literature is that we analyse differences in both assessment measures for track allocation and switching, where we are able to observe children's later outcomes in 9th grade in the form of track allocation, track switches and their scores on a math and language test.

Other studies have shown that test scores in secondary school are predictive of labour-market outcomes (e.g., [Murnane et al. 1995](#); [Currie and Thomas 1999](#)). We obtain a set of estimates suggesting that switching seems to lead to lower test scores in 9th grade. This seems to support arguments that switching tracks harms children's accumulation of human capital, which is documented in [van Elk et al. \(2009\)](#) and [Diris \(2012\)](#) for the Netherlands.

Our work also contributes to the literature on early school tracking. The long-run effects of early tracking for human capital development and educational opportunities have been summarized by [Hanushek and Woessmann \(2006\)](#) and [Brunello and Checchi \(2007\)](#). According to the OECD, the early tracking regime in the Netherlands causes a severe constraint for the growth of higher education participation. It states that "In the end, postponement of the present early tracking regime seems inevitable; although this is a major change in the way Dutch society thinks of itself" ([OECD 2007](#), p. 38). Consistent with this advice, other studies using Dutch data suggest that relatively low-

⁴ We group the tracks in pre-vocational education which leads to eight educational tracks as the number of children allocated to each of these tracks is relatively low. This decreases the variation in the data compared to analyzing 11 educational tracks.

ability children could improve their educational outcomes by about 30 % points when tracking is postponed by one year. At the same time, most children do not seem to be hurt by the presence of low-ability children in the first year of secondary school. Only children who are considered to have the highest ability seem to be hurt by the presence of lower ability peers (e.g., [van Elk et al. 2009](#); [Diris 2012](#)). We show that a substantial fraction of our population is not allocated to the right track, which seems akin to restraints on optimal human capital development.

We proceed as follows. First, we present background features of the Dutch education system and explore the research strategy. Section 3 documents the data description and statistics of our core variables. Sections 4–6 present the results on the differences between objective and subjective assessment measures, track switching and 9th grade test scores. Section 7 briefly addresses the policy perspective of our analysis with a focus on reducing switching. Section 7 concludes. We present additional results and detailed data descriptions in the online appendix to this paper.

2 Background and Strategy

We observe five main outcomes for each child: the test score which serves as an objective assessment of ability at the end of primary school (6th grade), the primary school teacher's assessment which serves as a subjective assessment measure of ability (6th grade), track allocation in the first and third year of secondary education (7th and 9th grade), the results from a cognitive test in 9th grade, and track switching in the first three years of secondary education (7th–9th grade). We now present information on these measures and information about the Dutch education system.

2.1 Dutch Education System

Countries differ in the age at which they first track children into different types of schools. In the majority of OECD countries, tracking takes place between the ages of 14–16. Some countries, including the Netherlands, undertake the first tracking at the age of 12 when children progress from primary to secondary school (i.e., from 6th to 7th grade).⁵ We take advantage of this system by studying the allocation in secondary school, the transition from primary to secondary school and performance in 9th grade.

⁵ Research has been conducted to investigate the effects of early tracking on differences in human capital formation and educational opportunities (e.g., [Hanushek and Woessmann 2006](#); [Brunello and Checchi 2007](#) for comprehensive review studies). The overall conclusion of these investigations has been that the earlier the tracking, the greater the impact of socio-economic factors, such as parental education, on educational outcomes and opportunities. For the Netherlands, [van Elk et al. \(2009\)](#) and [Diris \(2012\)](#) find that early tracking reduces participation in and completion of education for relatively low-ability children placed in lower tracks at the age of 12, while there does not seem to be a positive effect on those placed in the higher tracks. Furthermore, [Korthals and Dronkers \(2016\)](#) argue that in countries with highly differentiated tracking systems, such as the Netherlands, students perform best when track placement is based on prior performance. Our paper investigates these prior measures of performance. One of the main reasons of implementing an educational system with early tracking is to provide children with a learning environment in which they benefit from classrooms with homogenous populations. Such populations would benefit from peer effects. In addition, it allows for a focused curriculum and appropriately paced instructions.

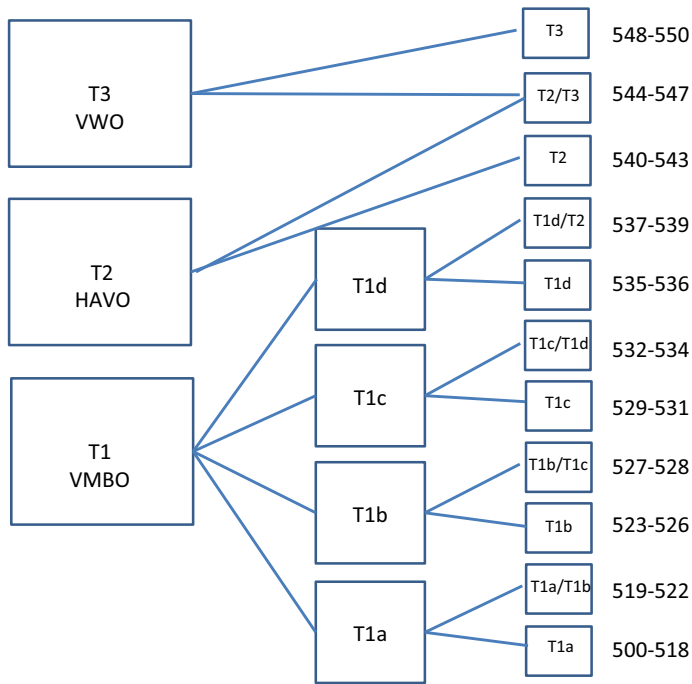


Fig. 1 Tracks in secondary education in The Netherlands. *Note:* The left-hand side shows the three major tracks from high (T3) to low (T1). The T1 track is subdivided into four sub-tracks. The right-hand side of the figure shows the 11 tracks to which children can be allocated in 7th grade and the objective assessment measure (test score) in the brackets that belongs to each of these tracks

In the Netherlands primary education consists of eight years of which the first two are spent in kindergarten. As of the third year of primary school (i.e., 1st grade), children formally learn how to read and write. Most children start kindergarten at the age of 4, enter 1st grade at the age of 6 and finish primary school at the age of 12. As of secondary school children are allocated to tracks. The track allocation decision is made by the secondary school. It is based on test scores and the teacher's assessment. Some schools set a threshold test score level below which children are not allowed to enter a certain level of secondary education.

The Dutch secondary education system is hierarchically structured by ability and consists of three main tracks that differ in duration and qualification (see the left-hand column of Fig. 1). The four-year track (VMBO or T1) qualifies children for vocational education, the five-year track (HAVO or T2) qualifies children for higher vocational education and the six-year track (VVO or T3) qualifies children for university education. The next column in Fig. 1 shows four sub-tracks at the lowest level of secondary education (T1a–T1d). The difference between these four sub-tracks is

the importance of a practical versus theoretical focus in the curriculum. Time spent on more theoretically oriented courses increases with the tracks from T1a to T1d.⁶

The third column of Fig. 1 shows all possible tracks, some of which are combinations of the three major tracks. Both the objective and subjective assessment measure are tailored toward allocating children into one of these 11 track combinations. The 6th grade test distinguishes brackets which are consistent with these 11 track combinations, shown in the fourth column of Fig. 1. Teacher assessment is measured on the same 11-point scale.

2.2 Background

In 6th grade, all children have to take an objective assessment test. Schools are free to choose which test their children take. Approximately 85 % of all Dutch children complete the *Cito Eindtoets*. We use results from this *Cito Eindtoets* for our analysis. The children in our data have taken this test in 2009. The test is standardized, meaning that the test procedure is the same for the whole country. During the assessment children have to answer questions in the areas of math, reading, study skills and science. The performance is measured on a scale between 501 and 550.

The aim of the cognitive test score is to provide an independent and appropriate perspective on children's expected performance and their best track placement in secondary education. The test institute offers guidelines for children's track allocation by reporting brackets of scores and accompanying track assessments. We followed these guidelines for constructing our objective ability measure variable. We use the brackets as the outcome measure of the objective assessment measure.

High scores on the standardized test are an important way in which primary schools try to signal the quality and value-added of their educational efforts. Primary schools seem to use their average scores on this test to attract new children. In addition, the Dutch Education Inspectorate uses these results, controlled for individual characteristics, as one of the inputs for their overall evaluation of the school's quality and value-added. Children also have an incentive to obtain a high test score because it is an important signal of their ability. In that sense, the test is a high-stakes assessment.

In addition to the objective test score, teachers make a personal assessment of each child's level of ability. The assessment is based on the teacher's experience and interaction with the child, observable demographic and socio-economic factors and the child's performance throughout all grades in primary school. Teachers also know the test score at the time they make their assessment of the child's ability. The subjective assessment is provided in the spring of 2009 before children apply to a secondary school. The teacher's assessment is provided in similar brackets as the objective assessment and fits with possible track allocations in secondary schools.

Primary school teachers do not have a strong incentive for strategic behaviour in such a way that their assessment overstates the child's ability. The teacher's compen-

⁶ The six-year track (T3) can also be divided into two sub-tracks of which one includes the option to take courses in Greek and Latin. Our data do not allow us to distinguish between these two sub-tracks, which means that we combine them in T3. In practice, test scores and teacher assessment do not distinguish the two tracks.

sation scheme does not depend on the assessments made. Furthermore, the primary school's population usually goes to the same secondary schools every year. This means that over time the information asymmetry between the primary and the secondary schools reduces and secondary schools learn how to interpret the assessments from primary schools. Furthermore, each year children are assigned to a class in which they are taught by one primary school teacher. This teacher is involved in teaching all subjects in primary school. Differences in test scores across different parts of the test are therefore unlikely to be driven by different teacher characteristics.

Secondary schools allocate children to tracks. They obtain the information about the objective and subjective assessment measures. Secondary schools have an incentive to allocate children to the track that matches their ability level. Inputs for the Education Inspectorate evaluation of secondary schools' performance include the percentage of children who graduate every year as well as the percentage of children who switch tracks. Allocating children to tracks that are too high (too low) leads to switching downward (upward) and would induce negative (positive) evaluations on this part of the performance assessment. Nevertheless, secondary schools also benefit from having more children in the higher tracks as this is beneficial for signalling the quality of the schools' education, which potentially helps to attract more children.

2.3 Strategy

Our analysis first focuses on the way in which both the objective and subjective assessment measures in 6th grade help to explain track placement in 7th grade. Both assessment measures aim to measure ability, but face the problem that the true underlying and unobserved level of ability is unknown. The test score (TS_i) is used as the objective assessment measure of the child's (i) ability. This test score depends on the child's true and unobserved ability (A_i), a vector of observed characteristics (X_i) and the primary school he attends (P_i).

In an ideal world $TS_i = A_i$. In practice this is not the case because TS_i is measured with noise and observed characteristics X_i are likely to influence the measurement of A_i by TS_i . The reason for adding school fixed effects (or dummies in our cross-sectional specifications) is that characteristics of primary schools can be related to test results of children in 6th grade, which can influence TS_i . Some of these school characteristics we cannot observe. Hence, we add school fixed effects (β_{P_i}) to the model.

The teacher's assessment (TA_j) is used as the subjective assessment measure of the child's ability. This measure includes the same ingredients plus the observed test score. The information about the child's test score influences the teacher's assessment of the child's ability. Because children are assigned to one teacher in the final year of primary education and most primary schools have only one 6th grade class, potential teacher effects are captured by school fixed effects. This is also the reason for indexing all variables with child i and for ignoring teacher j effects.

Finally, we observe the child's initial track placement (TP_i) in 7th grade (at secondary school) and thereafter allocation in 9th grade. The decision about initial placement is made by the secondary school and includes the objective and subjective

tive assessment measures. Adding secondary school fixed effects to the model would create additional endogeneity as not all schools offer the same track levels and secondary school fixed effects are related to TS_i and TA_i .

In the first part of the empirical analysis we analyse whether there are any systematic differences between the test score, the teacher's assessment and track placement for various socio-economic background characteristics of children. We are primarily interested in explaining track placement. We estimate equation (1) with an ordered probit model to find the determinants of track placement in both 7th and 9th grade. We do not observe A_i but two signals TS_i and TA_i . In the empirical analysis we incorporate the possibility that secondary schools take into account both TS_i and TA_i , although TA_i includes information about TS_i . In this way we have the two signals competing with each other. A statistically significant coefficient of TS_i would in all likelihood suggest that secondary schools put weight on both signals of ability. The equation we first estimate is:

$$TP_i = C_1 + a_1 X_i + a_2 TS_i + a_3 TA_i + \beta p_i + \varepsilon_i, \quad (1)$$

where ε_i is the error term.

In the second part of the empirical analysis we estimate the determinants of track switching in the first three years of secondary school (i.e., in the period spanning 7th, 8th and 9th grade). To do so, we first estimate a set of probit models in which we show what type of children tend to switch tracks. Second, we estimate probit models in which we estimate the probability of switching tracks (SW_i) for child i :

$$SW_i = C_2 + b_1 X_i + b_2 TS_i + b_3 TA_i + b_4 TP_i + \gamma S_i + \mu_i, \quad (2)$$

where μ_i is the error term. We estimate different versions of the model in which the dependent variable is switch, switch up or switch down. We only include secondary school fixed effects (γS_i) because switching takes place in secondary school. We show below that primary school fixed effects have no impact on track placement, which makes us confident that including only secondary school fixed effects is sufficient to estimate the determinants of switching.

Finally, we estimate models to investigate to what extent test scores on an identical (low-stakes) test in 9th grade are correlated with the ability signals from the teacher and the test score in 6th grade. We also estimate to what extent switching is correlated with test scores in 9th grade.

The strength of the data at our disposal is that we are able to observe performance in both primary and secondary school. In addition, we have detailed information about teacher assessments and initial track placement in secondary school. This is a unique feature in the literature. Nevertheless, the analysis is constrained by the fact that we are not able to identify a source of exogenous variation in our data. Ideally, one would want to conduct an experiment in which a random portion of the sample was placed according to the test scores' signal, another part according to the teacher's signal and a final slice of the population as it is currently done (i.e., decided by the secondary school). The alternative is to find instruments to deal with the "self-fulfilling prophecy" that creates endogeneity. The "self-fulfilling prophecy" is the idea that when a child

is placed on a higher or lower track than he should be according to his true ability, the child is more likely to switch back to the track that matches his true ability. This is to be kept in mind when interpreting the results in the switching section. Furthermore, all tests, but also the teacher's assessment, contain measurement error and analyses concerning ability suffer from omitted variable bias. However, such instruments are not readily available. Our analyses focus on outcomes between 6th and 9th grade. In order to find exogenous variation we would need a set of instruments related to (one of) our assessment measures and at the same time unrelated to any unobservable variables that might influence our outcome variable. Since children's true ability is unobserved, this is problematic. We are aware of the endogeneity concerns with respect to omitted variable bias and also with the fact that potential measurement error is an important disclaimer when interpreting the estimated coefficients, but try to deal with this in the best way possible by using primary and secondary school fixed effects and a rich set of covariates, including track placement in 7th grade.

3 Data

Before we present our results, we first document the most salient features of our data to reveal information about the allocation of children and to present a number of key descriptive statistics. More information as well as additional regression analyses are presented in the online appendix to this paper.

3.1 Descriptive Statistics

For the analyses we use a unique dataset on the educational development of children in a given region (Limburg) of the Netherlands. These data are collected in a cooperative project between (primary and secondary) schools, school boards, municipalities and Maastricht University to analyse school performance in order to foster educational improvement. The unique feature of this project is the participation of almost all schools in the region, implying almost full coverage of children (about 95 % of the primary schools participate and about 90 % of the secondary schools in the region). In 2009, information about all 6th grade children was collected and these children were again reviewed when they were in 9th grade in 2012. In both years the data collection includes administrative data from the school information systems, surveys among children and their parents and test results. The data covers children from all tracks with the exception of those who are in special needs education.

Table 1 documents the distribution of test scores TS_i in 7th and 9th grade, teacher assessment TA_i , and (initial) track placement TP_i , across the 11 different tracks (see Fig. 1). The numbers in each row add up to 100 %. In addition, the table documents the number of switchers from the initial track to which they are allocated in 7th grade. The numbers represent the fraction of children who switch away from each of these tracks. The distribution of test scores and teacher assessment seems to be broadly consistent but there are also important differences. Track placement and teacher assessment at higher levels of secondary education is different from the test scores, with almost a quarter of the sample being placed in T3 in 7th grade. Teacher assessment seems to

Table 1 The distribution of test scores, teacher assessment, track placement and switching

	Track										
	T1a	T1a/b	T1b	T1b/c	T1c	T1c/d	T1d	T1d/2	T2	T2/3	T3
TS_i	2.00	2.62	3.87	2.76	6.33	9.07	7.11	13.98	20.49	18.96	12.82
TA_i	4.00	3.27	4.53	0.31	0.11	5.18	10.96	12.33	18.02	16.62	24.67
TP_i	2.69	3.11	7.27	0.00	1.53	1.84	10.73	10.84	11.07	25.93	24.98
SW_i	3.59	0.21	11.30	0.00	7.29	5.28	19.01	10.35	12.99	9.93	20.06
TS_i 9th	1.21	2.21	4.14	0.00	1.88	1.38	11.70	11.31	9.82	29.14	27.12

All rows add up to 100. $n = 4500$ except for SW_i where $n = 4019$ and TS_i 9th where $n = 1812$. See Fig. 1 for a schedule of the different educational tracks. TS_i is the objective assessment measure based on the *Cito Eindtoets* (test score) in 6th grade. TA_i is the subjective assessment measure based on the teacher advice in 6th grade. TP_i is track placement in 7th grade. SW_i is a measure of switching. The numbers in this row show the fraction of children that switch away from this track after initial placement in 7th grade. TS_i 9th is the test score in 9th grade on math or language

be more favourable for the higher tracks compared to the test score. We explore these differences further in Sect. 4.

Furthermore, teachers seem to be reluctant to advise tracks in the middle tracks of vocational education (i.e., tracks in T1). As can be observed from Table 1, some T1 sub-tracks contain only few observations. These are combination tracks to which only a few children are allocated.⁷ We merge the combination tracks and the regular tracks for T1 when analysing differences between TS_i and TA_i . This means that for the analysis of track allocation we merge tracks T1a and T1a/T1b, T1b and T1b/T1c and T1c and T1c/T1d into three categories. This results in a more conservative estimate of the determinants of allocation in Sect. 4. As the difference between TS_i and TA_i does not have a direct impact on the number of children who switch tracks, we use 11 tracks for the analyses of switching tracks. We define differences in TS_i , TA_i and TP_i when there is a difference of at least two tracks. Furthermore, we define a switch between tracks when children switch at least two tracks. For example, if children switch from track T2 to track T1d/T2 or track T2/T3 this is not defined as a switch. When children switch from track T2 to track T1d or track T3 this is defined as a switch.

Table 2 reports descriptive statistics. In 6th grade children are on average 12 years old. The majority of children (and their parents) were born in the region (Limburg). The average parental education level suggests that they have completed vocational education. Almost all fathers are employed and work fulltime. Mothers more often report to be unemployed, or work part-time. In The Netherlands, part time employment is an important form of employment for women with young children (e.g., Bosch et al. 2010). Almost 80 % of the children in our sample live with both parents in 9th grade.⁸

Teacher assessment, track placement and test score (short) are all measured on a scale from 1 to 8. The original test score in 6th grade is measured on a scale from 501 to 550. Based on test score ranges provided by the institution that supplies the

⁷ It is possible that this is related to the availability of this type of school track in the area.

⁸ The corresponding survey question is: "Do you live at home with both parents?"

Table 2 Descriptive statistics of the main variables in the empirical analysis

	N	Mean	SD	Min	Median	Max
Female	4422	0.51	0.50	0	1	1
Age (years)	4500	12.06	0.55	10	12	14.75
<i>Region of birth child</i>						
1. Limburg		89.1 %				
2. The Netherlands		6.7 %				
3. Abroad		4.2 %				
	3943	1.15	0.46	1	1	3
Region of birth father	3923	1.36	0.67	1	1	3
Region of birth mother	3935	1.35	0.67	1	1	3
<i>Education father</i>						
1. Lower education		26.2 %				
2. Vocational education		32.1 %				
3. Higher education		18.9 %				
4. University		22.8 %				
	3752	2.38	1.10	1	2	4
Education mother	3763	2.22	1.03	1	2	4
<i>Labour market position father</i>						
1. Employed		93.8 %				
2. Unemployed		2.0 %				
3. Sick/unable to work		2.5 %				
4. Other		1.7 %				
	3858	1.12	0.51	1	1	4
Labour market position mother	3869	1.48	1.04	1	1	4
<i>Workdays per week father</i>						
1. Part-time 1–2 days		0.2 %				
2. Part-time 3–4 days		8.3 %				
3. Fulltime		91.5 %				
	3249	2.91	0.29	1	3	3
Workdays per week mother	2848	2.31	0.62	1	2	3
Living with both parents	3608	0.79	0.41	0	1	1
Teacher assessment	4500	5.6	2.15	1	6	8
Track placement	4500	5.75	2.16	1	7	8

test in 6th grade, we rescaled the test score to a scale from 1 to 8. Average test score and average test score (short) correspond to a T1d/T2 track (vmbo-t/havo). Average teacher assessment and average track placement correspond to a T2 track (havo). Almost 24 % of children switch tracks between 7th and 9th grade. The majority of the switches happen in the pre-vocational track.

Finally, we use information about a math and language test that children in our sample have taken in 9th grade. This test was a low-stakes test and part of the research

Table 2 continued

	N	Mean	SD	Min	Median	Max
Test score	4500	538.67	8.03	504	540	550
Test score (short)	4500	5.21	2.02	1	6	8
Switching	4019	0.24	0.42	0	0	1

The same description of region of birth of the child applies to the father and mother. Similarly, the same description of education father, labor market position father and workdays per week father apply to the mother. The parental education level is measured as the highest completed level of education and measured in four different categories: (1) lower education, (2) vocational education, (3) higher education and (4) university education. The parental labor-market position is measured in four categories: (1) employed, (2) unemployed, (3) sick/unable to work and (4) not in the labor force. The number of workdays per week of the parents is a measure of the amount of time allocated to paid work. Working time is defined in three categories: (1) part-time 1–2 days per week, (2) part-time 3–4 days per week and (3) fulltime. The corresponding survey question for living with both parents is: “Do you live at home with both parents?”

project conducted at the schools in our sample. Its main purpose was to have a school-independent test score for students in 9th grade. The difficulty level of the test differs according to the students’ track level. Since we want to compare the effect of switching tracks across tracks we only use the questions on math or language that are identical for all students.⁹ We observe that the correlations among teacher assessment, test scores and track placement are high and positive.¹⁰ Interestingly, track placement shows a higher positive correlation with teacher assessment than with test scores. Furthermore, we observe a negative correlation between switching tracks and the test score, teacher assessment and track placement.

3.2 Possible Selection

Data has been collected from 155 primary schools (95 % of all schools in the given region) and 30 secondary schools (90 % of all schools in the given region). This results in a database of $n = 4500$ for the first section, $n = 4019$ for the second section and $n = 1812$ for the third section of the empirical analyses of this paper. In the first section of the empirical analyses we discuss non-random differences between the test score, the teacher assessment and track placement in 7th and 9th grade. In the second section of the empirical analyses we discuss track switching because mistakes in initial track placement or suboptimal allocation can be made undone in the first part of secondary education. In the third section we correlate both ability signals and switching with 9th grade test scores. Data has been collected for 9092 children. For the analyses in the

⁹ This leaves us with 11 questions on math and 8 questions on language. Since we do not have enough observations for all tracks, we only look at students in track T1c, T1d, T2 and T3 in relation to their score on the math or language test. The average percentage of questions answered correctly increases from 34 in track T1c to 49 in track T3. The standard deviation is around 20 % in all tracks. The distribution of the test score in 9th grade is presented in the last row of Table 1. The row adds up to 100 % and shows the percentage of students per initial track placement in 7th grade.

¹⁰ A full correlation table of all variables of interest and a detailed description of the covariates can be found in the online appendix.

first section of the paper we only use those children for whom we observed the teacher assessment and the test score in our data. For the analyses in the second section we also need to know children's track placement in 7th, 8th and 9th grade, which reduces the sample size to 4019. Finally, 9th grade test scores are available for 1812 children.

We address possible selection issues with regard to the sample we use for our analyses. We investigate whether individual characteristics are able to predict whether children end up in our sample for analyses. We are mainly concerned about schools only reporting data for their well performing children and holding back information on other children. After controlling for school fixed effects we find that individual characteristics are not significantly related to selection into our sample. We conclude that selection is not an issue (see the online appendix B for a more elaborate analysis).

4 Track Placement

This section presents our first set of estimation results. We first investigate to what extent there are non-random differences between the objective and subjective assessment measures. Second, we present a set of estimates about the determinants of track placement in 7th and 9th grade.

4.1 Differences Between Objective and Subjective Assessment Measures

To compare differences between our objective and subjective assessment measures we have created three categories: $TA_i < TS_i$, $TA_i = TS_i$ and $TA_i > TS_i$. About 81 % of the children in our sample is faced with objective and subjective assessment scores that are equal. If there is no systematic difference between the teacher assessment and the test score, we would expect both assessments to be equal on average. Any deviations should be approximately symmetric. However, we observe that it is more likely that the subjective assessment measure is higher when there is a difference between the two measures. In 5.1 % of the cases the subjective assessment is lower compared to the objective assessment, in 13.9 % it is higher. Since the teacher assessment makes use of the information revealed by the test score, the teacher has access to a child's educational history in primary school and has knowledge about a child's background characteristics and earlier test scores, she has an information advantage.¹¹

When we look into the differences between the test score, teacher assessment and track placement in 7th grade, we find that most of the differences we observe are related to gender and social-economic status.¹² First, we observe that girls seem more likely to receive a teacher assessment that is higher compared to the test score, while they seem less likely to receive a teacher assessment that is lower compared to the test score. In addition, girls are more likely to receive a track placement equal to the subjective assessment measure and more likely to have a test score that is lower com-

¹¹ If we use 11 different tracks instead of 8, 32.9 % of all children is faced with different assessment measures. Of these children 83 % receives a higher subjective assessment measure of ability.

¹² Table C1 and C2 in the online appendix show descriptive statistics and estimated marginal effects from ordered probit models.

pared to their track placement. This suggests that girls not only get a more favourable assessment from the primary school teacher, but also with regard to track placement in secondary school. Second, from the labour market position of the mother we observe that children who have mothers who are unemployed are less likely to receive a subjective assessment that is higher than the objective assessment. These children are also less likely to have a test score that is lower than their initial track placement and more likely to receive a track placement that is equal to their test score. Furthermore, children of mothers who are sick or unable to work are less likely to receive a subjective assessment that is higher than the objective assessment. These children are also less likely to receive a test score that is lower than their track placement and more likely to receive a track placement that is equal to or lower than their objective assessment. This indicates that secondary schools seem to allocate these students unfavourably. In our data over 70 % of the mothers who are unemployed and over 85 % of the mothers who are sick or unable to work completed only lower education or vocational education. [Burgess and Greaves \(2013\)](#) obtain similar results with respect to ethnic minorities, which they attribute to negative stereotyping of particular groups in society.

4.2 Determinants of Track Placement

Figure 2 shows how children are allocated to tracks according to the two assessment measures. The figure is divided into three panels. Panel A documents track placement of children who are faced with $TA_i < TS_i$. Panel B displays track placement of those with $TA_i = TS_i$ and panel C shows placement of those with $TA_i > TS_i$. Placement in the category labelled “else” represents those children who are placed in a track that was recommended by neither the subjective assessment nor the objective assessment measure. In almost all of these cases the subjective and objective assessment measures differ by more than three levels and track placement is in between these two measures. In some cases track placement is higher or lower than both assessment measures indicate. In the latter case our data suggest that it is more likely that track placement is higher than both assessment measures would merit.

The bars in Panel A of Fig. 2 suggest that when $TA_i < TS_i$ children are more likely to be placed according to TA_i than TS_i (40.2 vs. 17.9 %). At first sight, this suggests that secondary schools seem to act in a relatively conservative way by following the lower of the two signals. They seem to attach more value to the teacher’s assessment of the child’s ability relative to the test score. At the same time, the share of children placed in tracks that do not directly correspond with one of the assessment measures is relatively large (41.9 %). The numbers in Panel C suggest that secondary schools are more likely to follow the teacher’s assessment even when it is higher than the test score. More than two thirds of the population with $TA_i > TS_i$ is placed according to TA_i . Also the share of children placed in tracks that do not directly correspond with the assessment measures is relatively low compared to the case in which $TA_i < TS_i$. Combined with the information from Panel A, a picture emerges that secondary schools attach more value to the teacher’s assessment of children’s ability relative to test scores. Note that the teacher’s assessment of a child’s ability is on average higher than the test score would suggest. Finally, the statistics in Panel B of Fig. 2 reveal that when

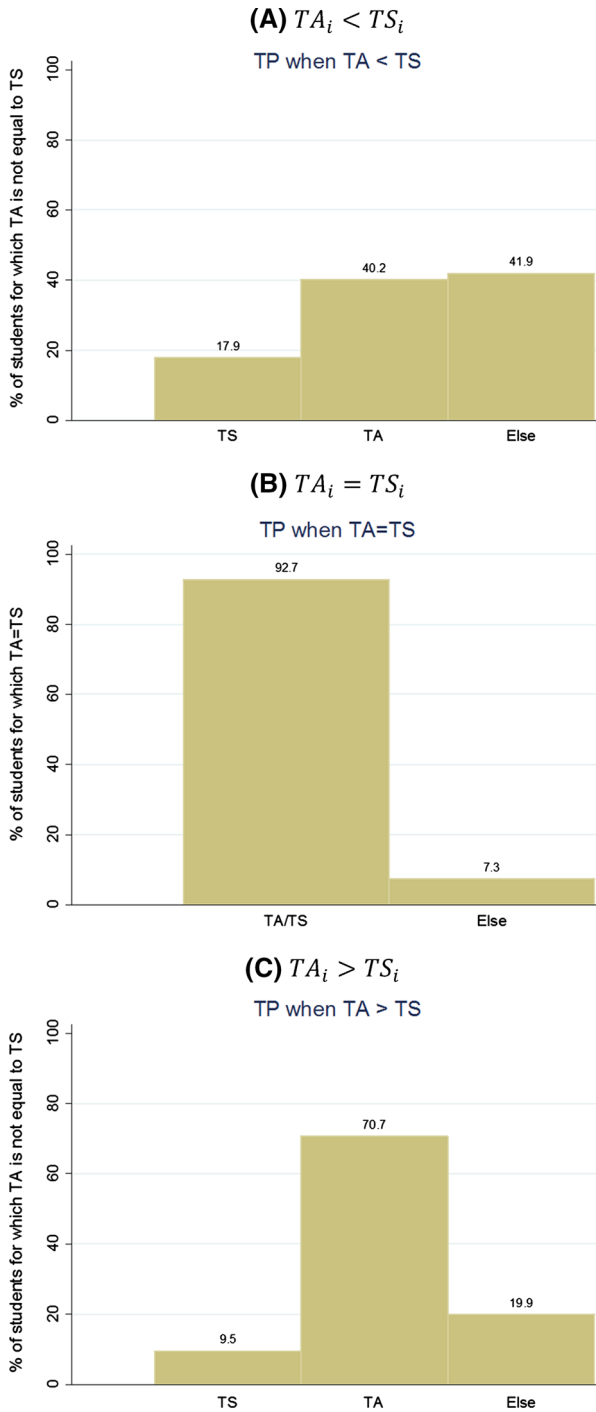


Fig. 2 Track placement for different objective and subjective assessment measures

both assessment measures give the same signal about children's ability almost all children are placed in the corresponding track. Nevertheless, 7.3 % of the children are allocated to a different track. Upon closer inspection most of these children are allocated to higher tracks relative to what the teacher's assessment and the test score recommend. Overall, it seems to be the case that secondary schools have a preference to allocate children according to the teacher's assessment measure and/or according to the assessment that signals the highest ability.

We continue by estimating the determinants of track placement in 7th grade. Table 3 presents the estimation results of Eq. (1). The estimates are coefficients from ordered probit models where track placement (measured between 1 and 8) is the dependent variable. In the first two columns we either use the test score or the teacher assessment to estimate the determinants of track placement. As expected, we observe a strong positive relation between both the test score and track placement and teacher assessment and track placement. When we rescale the test score and the teacher assessment by the cut points in their respective regressions we observe that a one standard deviation increase in the test score (teacher assessment) bridges 35.1 % (35.2 %) of the gap between the lowest and the highest track placement, without adding any other control variables. In column (3) we add both the test score and the teacher assessment to the model. The test score and teacher assessment do not seem to be orthogonal. When the test score and teacher assessment are both added to the model we observe that a one standard deviation increase in the test score (teacher assessment) bridges 11.7 % (25.2 %) of the gap between the lowest and the highest track placement, without adding any other control variables. The difference in coefficients between the test score and teacher assessment is significant. An important observation is that teacher assessment seems to be a better predictor of track placement than the test score. In columns (4)–(6) we add control variables, primary school fixed effects and a measure of children's GPA on math and language tests in previous years, respectively. Our estimates remain approximately the same in these different specifications. Overall, both teacher assessment and the test score seem likely to be important determinants of track placement in 7th grade. From our final specification in column (6) we conclude that teacher assessment appears to play a more important role in determining track placement compared to the test score. The estimated coefficient is about twice as large. This finding seems to be consistent with the subjective assessment measure having more information about the child's ability than the objective assessment measure.

In columns (7)–(12) we investigate the determinants of track placement in 9th grade. Track placement in 9th grade consists of six categories as over time the combination tracks disappear and children get allocated to their final track.¹³ Columns (7)–(9) show the results for the sample of children who have not switched tracks between 7th and 9th grade. We obtain estimated coefficients that suggest that teacher assessment is still the best predictor of track placement for the children who have not switched tracks.

¹³ Secondary schools use combination tracks in the first year(s) of secondary education due to the uncertainty around ability. By using combination tracks schools can observe performance and allocate children to the most appropriate track after learning more about the ability. The combination tracks disappear into the neighboring main tracks. For example, children placed in track T2/T3 in the first year are expected to be re-placed in track T2 or T3 in year 2 or 3. Such a re-placement is not considered as switching tracks.

Table 3 The determinants of track placement (dependent variable TP_i)

Track placement														
7th grade			9th grade											
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)			
Test score (TS)			1.517*** (0.093)	0.687*** (0.106)	0.762*** (0.087)	1.102*** (0.142)	0.980*** (0.263)	1.007*** (0.072)	1.501*** (0.115)	1.285*** (0.280)	0.854*** (0.076)	1.037*** (0.110)	1.086*** (0.199)	
Teacher assessment (TA)			1.998*** (0.246)	1.490*** (0.249)	1.463*** (0.246)	1.847*** (0.279)	1.941*** (0.278)	1.690*** (0.206)	2.173*** (0.209)	2.097*** (0.273)	1.030*** (0.249)	1.612*** (0.183)	1.481*** (0.204)	
Cut1			-2.994***	-4.029***	-4.162***	-4.255***	-5.875***	-11.266***	-6.168***	-9.515***	-11.529***	-4.152***	-7.459***	-11.350***
Cut2			-2.027***	-2.410***	-2.553***	-2.796***	-4.115***	-9.060***	-3.564***	-6.732***	-8.574***	-2.863***	-5.378***	-8.905***
Cut3			-1.712***	-1.890***	-2.031***	-2.011***	-3.419***	-8.408***	-3.009***	-5.391***	-7.637***	-2.181***	-3.968***	-7.841***
Cut4			-0.962***	-0.889***	-0.994***	-0.811	-1.895***	-7.245***	-1.246**	-3.476***	-5.660***	-0.705	-2.703***	-6.233***
Cut5			-0.365*	-0.160	-0.216	0.062	-0.523	-5.483***	1.001*	-0.601	-2.575***	0.785**	-0.629	-3.885***
Cut6			0.151	0.432**	0.414**	0.651	0.323	-4.719***						
Cut7			1.322***	1.651***	1.734***	2.199***	1.720**	-3.319**						
Test score/(C5-C1)			0.351		0.117	0.118	0.145	0.123	0.140	0.168	0.144	0.173	0.152	0.145
Teacher assessment/(C5-C1)			0.352		0.253	0.227	0.243	0.244	0.236	0.244	0.234	0.209	0.236	0.198
Difference TS-TA (<i>p</i> value)					0.012	0.011	0.030	0.025	0.001	0.002	0.056	0.562	0.018	0.229
Controls			No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Primary school FE			No	No	No	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes
GPA1-1,2,3			No	No	No	No	No	Yes	No	No	Yes	No	No	Yes
Observations			4,500	4,500	4,500	2,509	1,100	572	1,708	872	469	2,509	1,100	572

Coefficients are from ordered probit regressions with track placement as the dependent variable. The dependent variable in columns 1–6 is track placement in 7th grade for all children. The dependent variable in columns 7–12 is track placement in 9th grade, where columns 7–9 are for children who have not switched tracks and columns 10–12 are for all children. The dependent variable in 9th grade consists only of 6 tracks as over time the combination tracks disappear and children get allocated to a final track. Standard errors are clustered at the secondary school level. ***, **, * represent 1, 5 and 10 % significance levels respectively. TS and TA are standardized. The control variables are gender, age, birth region of the child, birth region of father and mother, education level of father and mother, labour market position of father and mother, work days per week of father and mother, if the child lives with both parents, area code of residence and region. $GPA_{1-12,3}$ is the child's GPA in primary school in previous years

Columns (10)–(12) show the results for the entire sample of children we observe in 9th grade. We obtain a set of coefficients that suggests that there is no statistically significant difference between the predictive power of the test score and teacher assessment. These results indicate that the teacher is the best predictor of children's ability both in 7th grade and later on in the children's secondary school career. However, the results also show that the predictive power of the teacher compared to the predictive power of the test score falls over time. A possible explanation for this is that children who were initially assessed too favourable by the teacher have switched to a lower track in the first three years of secondary education.

To take into account the covariation between the teacher assessment and the test score we also estimate the predictive power of the teacher assessment on track placement in 7th and 9th grade after correcting for the predictive power of the test score. This seems a natural thing to do because the teacher knows the test score of the child when the assessment is made. We find that the teacher's assessment is still highly predictive of track placement in 7th and 9th grade after controlling for the covariation between the teacher's assessment and the test score. The estimated coefficients of these analyses can be found in Appendix E.

5 Switching Tracks

Children could be allocated sub-optimally across different levels of secondary education. Suboptimal allocation encourages switching, which comes with a cost of suboptimal human capital investments and adjustment cost. In addition, children could have to stay for an additional period in the secondary education system because transitions between tracks are not always smooth. Note that switching tracks is the most drastic measure that secondary schools can take when children are not performing up to expectations. For example, the first option for children who are not able to keep up with the required level is to let them repeat the same grade. In the event that the school believes that grade retention will be insufficient, children have to switch tracks. Furthermore, it is possible that the costs of switching tracks are different when children switch up tracks compared to when they switch down tracks. It is likely that the costs of switching tracks for children who switch down are more related to demotivation and that the cost of switching up tracks are more related to previous underinvestment of human capital.

5.1 Documenting Switchers

Figure 3 shows that approximately 24 % of children switches tracks between 7th and 9th grade. Most of the switches (55 %) happen from 8th to 9th grade.¹⁴ Approximately 71 % of all children who switch between tracks switch down and only about 29 % of all children who switch between tracks switch up. Switches are defined for all 11 tracks and are counted based on major switches, i.e., at least two steps. We observe that most

¹⁴ The online appendix D documents the characteristics of switchers in detail and presents the coefficients of probit models.

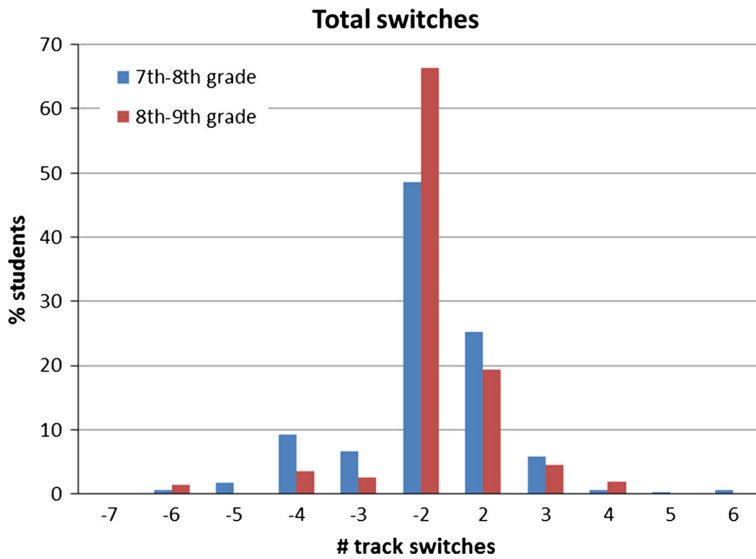


Fig. 3 Track switching between initial track allocation in 7th grade and 9th grade. *Note:* The figure shows the total number of switchers. Negative numbers on the horizontal axis are defined as switches down and positive numbers as switches up. The blue and red bars add up to 100 % individually. (Color figure online)

children who switch tracks switch two tracks up or down.¹⁵ Furthermore, correlations between teacher assessment, test score, track placement and switching are negative. This is due to the nature of our data. Since we observe track switches until 9th grade, we capture almost all of the switches in the T1 track but we capture less of the switches in the T2 and T3 tracks as these children still have the opportunity to switch tracks after 9th grade. This finding is confirmed by results from the Inspectorate of Education for all children in The Netherlands (Education Inspectorate 2007).

Figure 4 shows track switching by differences in objective and subjective assessment measures. The figure is divided into three panels. The first panel documents the track switching of children who are faced with $TA_i < TS_i$. The second panel displays track switching of those with $TA_i = TS_i$ and the final panel shows switching of those with $TA_i > TS_i$. Track placement is equal to either TA_i , TS_i or “else”. The category labelled “else” represents those children who are placed in different tracks than the measures advised. Note that the number of switchers is determined on the basis of all 11 possible tracks.

The bars in Panel A suggest that when $TA_i < TS_i$ fewer children switch between tracks when they have been placed according to TA_i and more children switch when they are placed according to TS_i or in another track than either measurement pointed at. Comparison of TA_i and TS_i in Panel A suggests that those who are placed according to TS_i are more likely to switch up consistent with the argumentation that the teacher

¹⁵ Our data includes 1266 switchers. Of these switchers 43 children switch down/up multiple tracks and 65 children switch both up and down between 7th and 9th grade. We analyze overall switches which means that we refer to 1158 switchers (i.e., $1266 - 43 - 65 = 1158$). Figure 3 includes all 1266 switchers.

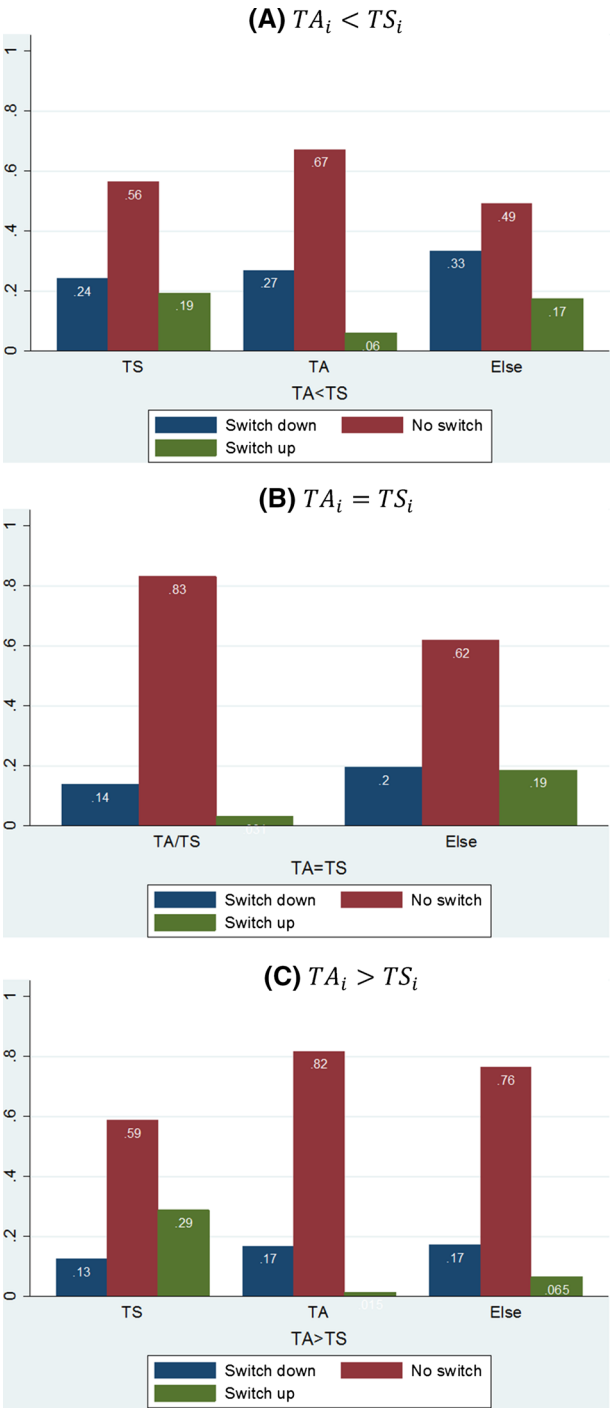


Fig. 4 Track switching for different objective and subjective assessment measures

Table 4 Switching between tracks

	Switch			Switch down			Switch up		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PL = TA \neq TS	0.099*** (0.029)	0.093*** (0.030)	0.098*** (0.032)	0.103*** (0.026)	0.091*** (0.025)	0.085*** (0.027)	-0.017 (0.012)	-0.010 (0.010)	-0.000 (0.007)
PL = TS \neq TA	0.319*** (0.101)	0.161*** (0.049)	0.179*** (0.078)	0.102*** (0.038)	0.024 (0.043)	0.025 (0.038)	0.252*** (0.117)	0.096*** (0.036)	0.084*** (0.042)
PL \neq TA \neq TS	0.269*** (0.100)	0.197*** (0.065)	0.167*** (0.081)	0.131*** (0.060)	0.090 (0.073)	0.079 (0.062)	0.164* (0.097)	0.105*** (0.038)	0.059* (0.035)
Controls	No	No	Yes	No	No	Yes	No	No	Yes
Secondary school FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	4019	3836	2499	4019	3664	2281	4019	3031	1750

Probit regressions with marginal effects. Standard errors are clustered at the secondary school level. ***, **, * at 1, 5 and 10% significance levels respectively. The base level in these regressions is when PL=TA=TS and when PL=TA and PL=TS but TA \neq TS. This can happen for example when TA = T3, TS = T2 and PL = T2/T3. In column 1, the base level represents 52.97% of the children. The three other categories shown in the table represent 30.28, 5.8 and 10.95% respectively. We control for gender, age, birth region of the child, birth region of the father and mother, education level of the father and mother, labour market position of father and mother, if the child lives with both parents, area code of residence and region

assessment is generally more generous about the children's ability than the test score. The numbers in Panel C suggest that although secondary schools are more likely to follow the teacher's assessment, even when it is higher than the test score, the number of switchers is relatively low when children are allocated based on TA_i . In addition, if children are placed according to TS_i more children switch up. Finally, the statistics in Panel B of Fig. 4 reveal that children switch less often if $TP_i = TA_i = TS_i$. The overall picture that emerges from Fig. 4 is that children placed according to TA_i have a lower probability to switch tracks relative to children placed according to TS_i . This conclusion seems to hold regardless of the difference between TA_i and TS_i . In addition, children allocated in accordance with the highest (lowest) of the two assessment measures seem to have a higher probability to switch down (up) a track, which is consistent with overassessment (underassessment) of a child's ability.

5.2 Determinants of Track Switching

We continue by presenting the results of analysing probit models in which we estimate the probability of switching. Columns (1)–(3) in Table 4 present marginal effects for overall switching, columns (4)–(6) present marginal effects for switching down and columns (7)–(9) present marginal effects for switching up. The third specification in all three models includes control variables and secondary school fixed effects. Furthermore, standard errors are clustered at the secondary school level.

The estimated coefficients in column (3) suggest that children who are placed in a track according to the teacher's assessment (and not according to the test score) are 9.8 % more likely to switch tracks, children who are placed according to the test score (and not according to the teacher assessment) are 17.9 % more likely to switch tracks and children who are placed not according to the teacher's assessment or the test score are 16.7 % more likely to switch tracks compared to the base level.¹⁶ The coefficients displayed in column (6) suggest that children placed in a track according to the teacher's assessment (and not according to their test score) are more likely to switch down, whereas children who are placed according to their test score (and not according to their teacher's assessment) or according to neither of the assessment measures are not more likely to switch down. Finally, from the estimated coefficients shown in column (9) we observe that children who are placed in a track according to their test score (and not according to their teacher's assessment) and children who are placed according to neither of the assessment measures are more likely to switch up, whereas children placed according to their teacher's assessment (and not according to their test score) are not more likely to switch down. Finally, children for whom $TP_i = TA_i = TS_i$ are least likely to switch tracks. For these children there is only little uncertainty about the most efficient track placement. These results seem to suggest that children are less likely to switch when they are allocated to tracks based on the teacher's assessment. However, the teacher's assessment is generally more favourable than the test score about a child's ability. Therefore, children who are allocated according to

¹⁶ The base level in all these analyses is when $TP_i = TA_i = TS_i$ or $TP_i = TA_i$ and $TP_i = TS_i$ but $TA_i \neq TS_i$. For example: $TA = T3$, $TS = T2$ and $PL = T2/T3$.

Table 5 Determinants of scores on math or language tests in 9th grade

	Math or language test 9th grade			
	(1)	(2)	(3)	(4)
Test score	0.027* (0.014)		0.006 (0.015)	0.002 (0.016)
Teacher advice		0.071*** (0.016)	0.067*** (0.018)	0.060*** (0.018)
Controls	Yes	Yes	Yes	Yes
Secondary school FE	No	No	No	Yes
Observations	1116	1116	1116	1116
Adjusted R-squared	0.072	0.082	0.082	0.085

Estimates are coefficients from OLS regressions. The dependent variable is the children's score on a math or language test in 9th grade. Test score and teacher advice are standardized. Standard errors are clustered at the secondary school level. ***, **, * at 1, 5 and 10% significance levels respectively. We control for track placement in 7th grade, gender, age, region of birth of child, father and mother, education level of father and mother, labour market position of father and mother, days worked of father and mother, if the child lives with both parents, area code of residence and region

the teacher's assessment are more likely to switch down and children allocated to the test score are more likely to switch up.¹⁷

6 Test Scores in 9th Grade

For a subsample of children we have data about a math and language test score in 9th grade. Children were randomly assigned questions in either math, language or both. We use the answers to 11 math questions or 8 language questions that have been asked to children in all tracks. This results in test scores of $n = 1812$ children.

Table 5 documents the estimated coefficients of an analysis in which we investigate the effect of teacher assessment and test scores in 6th grade on the test score in 9th grade. The coefficients shown are coefficients from OLS regressions in which the dependent variable is the child's score on the test in 9th grade. The test score and teacher assessment are standardized and standard errors are clustered at the secondary school level. Columns (1) and (2) show that both the test score and the teacher's assessment in 6th grade are positively and statistically significant related to children's test score in 9th grade. Furthermore, the estimated coefficients suggest that the test score and the teacher's assessment are not orthogonal. Put together, the teacher's assessment seems to be able to predict the test score in 9th grade more accurately.

¹⁷ In the online appendix F we look at the type of school children attend. We see that most of the track switching takes place within the pre-vocational education tracks. This is consistent with findings from the Dutch Education Inspectorate. Furthermore, in the switching analyses we control for the type of school (level of comprehensiveness) children attend.

Table 6 The impact of switching tracks on math and language tests in 9th grade

Test in 9th grade: math or language											
Pre-vocational education (GL)				Pre-vocational education (TL)				Pre-higher education			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Switch	-0.142* (0.045)	-0.528** (0.119)	-0.543** (0.123)	0.027 (0.030)	-0.000 (0.091)	0.023 (0.106)	-0.032 (0.020)	-0.097*** (0.029)	-0.102** (0.035)	-0.073 (0.043)	-0.240** (0.089)
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
Secondary school FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No
Observations	88	38	38	317	155	155	603	303	303	732	415
Adjusted R-squared	0.059	0.092	0.022	0.002	0.048	0.081	0.002	0.042	0.054	0.004	0.034
Switch down	-0.103** (0.018)	-0.528** (0.119)	-0.543** (0.123)	-0.001 (0.025)	-0.000 (0.091)	0.023 (0.106)	-0.039 (0.025)	-0.160** (0.060)	-0.162** (0.064)		
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes		
Secondary school FE	No	No	Yes	No	No	Yes	No	No	Yes		
Observations	88	38	38	317	155	155	603	303	303		
Adjusted R-squared	0.032	0.092	0.022	-0.003	0.014	0.078	0.003	0.054	0.064		
Switch up				0.044* (0.023)	-0.166** (0.055)	-0.204*** (0.040)	0.022 (0.045)	0.079 (0.071)	0.110 (0.084)	-0.073 (0.043)	-0.240** (0.089)
Controls				No	Yes	Yes	No	Yes	Yes	No	Yes
Secondary school FE				No	No	Yes	No	No	Yes	No	No
Observations				317	155	155	603	303	303	732	415
Adjusted R-squared				0.061	0.049	0.079	-0.001	0.031	0.051	0.004	0.034

Estimates are coefficients from OLS regressions. The dependent variable is the children's score on a math or language test in 9th grade. Standard errors are clustered at the secondary school level. ***, **, * at 1, 5 and 10 % significance levels respectively. We control for test score and teacher advice in 6th grade, track placement in 7th grade, gender, age, region of birth of child, father and mother, education level of father and mother, labour market position of father and mother, days worked of father and mother, if the child lives with both parents, area code of residence and region

This effect is robust to the inclusion of secondary school fixed effects, which suggests that this effect is not specific to certain (characteristics of) schools.¹⁸

We also estimate the predictive power of the teacher assessment for children's test scores in 9th grade after controlling for the covariation between the teacher's assessment and the test score in 6th grade. This analysis results in an estimated coefficient of 0.0395 that is statistically significant at the 10 % level. So, even after controlling for the covariation between the teacher's assessment and the test score, the assessment seems to be more predictive of later test scores compared to the 6th grade test score. It seems likely that the teacher's assessment also captures other skills, besides intelligence, that are important determinants of children's school career.¹⁹

Table 6 shows the results of an analysis in which we explain the effects of switching on 9th grade test scores. We compare children who have switched tracks to children who have not. The coefficients are from OLS regressions with the test score in 9th grade as the dependent variable. The standard errors are clustered at the secondary school level. Table 6 has three panels. The top panel presents estimates for overall switching, the middle panel presents estimates for switching down and the bottom panel presents estimates for switching up. In the top panel the coefficients for switching show that for children in the pre-vocational track (GL in Fig. 1), the pre-higher education track and the pre-university track there is a statistically significant negative effect of switching tracks on the 9th grade test score. Second, the coefficients displayed in the middle panel of the table suggest that children in the pre-vocational track (GL) and children in the pre-higher education track experience a statistically significant negative effect of switching down tracks. Finally, in the bottom panel of Table 6 the displayed coefficients suggest that children in the pre-vocational education track (TL) and children in the pre-university track experience a statistically significant negative effect of switching up tracks. Overall, it seems to be the case that children who have switched tracks experience a negative effect on their test score in 9th grade compared to children who have not switched tracks.²⁰

7 Conclusions

This paper documents and interprets the determinants of track placement in the transition from primary to secondary education and the first three years of secondary education. Our main findings suggest that both objective and subjective assessment measures of ability in 6th grade predict track placement in 7th grade. The subjective

¹⁸ The test in 9th grade is limited (8–11 questions) and is a low-stakes test. The correlation between children's test score in 6th grade and their test score in 9th grade is 0.16 and statistically significant at the 1 % level. For the analyses in Table 6 one could argue that the teacher assessment is better at predicting children's test score in 9th grade as the teacher assessment likely captures other elements next to children's ability that affect children test score on a low-stakes test more compared to their test score on a high-stakes test. The test score in 6th grade could be a better predictor of children's test score on a high-stakes test but such information is unavailable.

¹⁹ This analysis can be found in online Appendix E.

²⁰ Here we assume that track switches are caused by performance and not related to unobservable characteristics, such as illness or behavioral problems.

assessment measure of the teacher's assessment of the child's ability contains more information as the teacher has more knowledge about the child's socio-economic background, the objective test score in 6th grade and previous test scores and other results. Our estimates suggest that the teacher's assessment in primary school is a better predictor of a child's track placement and subsequent performance in secondary school compared to the 6th grade test score. We also observe that approximately 24 % of our population of children switches tracks between 7th and 9th grade. We obtain a set of estimates that suggests that children are the least likely to switch when they are allocated to a track based on the teacher's assessment. However, when we look at switching down and up separately our estimates suggest that children placed in tracks according to teacher assessment are more likely to switch down and children placed in tracks according to the test score and children allocated not according to any of these assessment measures are more likely to switch up. Finally, test scores in 9th grade seem to be better predicted by the teacher's assessment compared to 6th grade test scores. In addition, switchers obtain lower test scores on this 9th grade test relative to children who remain in their initial tracks.

This research uses a straightforward research design and explores a dataset which includes information on assessment measures, track placement and subsequent performance. Future work could extend our analysis by using for example more detailed information about different parts of the objective assessment measure. Test scores could be decomposed in a language and math component, which could benefit the analysis of allocation and performance in secondary school. In addition, the relationship with individual characteristics is interesting to explore further. We have used a limited number of covariates because of data limitations, but future data collection efforts also include measures of behaviour and personality traits. This information could help in estimating more precise coefficients and possibly additional determinants of performance and allocation. Finally, the children in our database are followed throughout the rest of their educational careers. This opens avenues for future research about longer term effects of track allocation.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bernardi, M., Bratti, M., & De Simone, G. (2014). *"I wish I knew..."—Misperceived ability, school track counselling services and performances in upper secondary education*, IZA Discussion Paper no. 7940. <http://ftp.iza.org/dp7940.pdf>.
- Bosch, N., Deelen, A., & Euwals, R. (2010). Is part-time employment here to stay? Working hours of Dutch women over successive generations. *Labour*, 24(1), 35–54.
- Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, 22(52), 781–861.
- Burgess, S., & Graves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31(3), 535–576.

- Cornwell, C., Mustard, D. B., & van Parys, J. (2012). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1), 236–264.
- Currie, J., & Thomas, D. (1999). *Early test scores, socioeconomic status and future outcomes*, NBER Working Paper no. 6943. <http://www.nber.org/papers/w6943>.
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity or gender matter? *American Economic Review*, 95(2), 158–165.
- Diris, R. (2012). *The economics of the school curriculum*, PhD thesis, Maastricht University.
- Education Inspectorate. (2007). *Transition from primary to secondary education*. http://www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2007/Aansluiting+vo+bo+printversie.pdf.
- Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, 104(8), 2567–2591.
- Gibbons, S., & Chevalier, A. (2008). Assessment and age 16+ education participation. *Research Papers in Education*, 23(2), 113–123.
- Golsteyn, B. H. H., & Schils, T. (2014). Gender gaps in primary school achievement: A decomposition into endowment and returns to IQ and non-cognitive factors. *Economics of Education Review*, 41, 176–187.
- Hanushek, E., & Woessmann, L. (2006). Does early tracking affect educational inequality and performance? Differences-in-differences evidence across countries. *Economic Journal*, 116(510), c63–c76.
- Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. Paris: OECD. <http://www.oecd.org/edu/ceri/Fostering-and-Measuring-Skills-Improving-Cognitive-and-Non-Cognitive-Skills-to-Promote-Lifetime-Success.pdf>.
- Korthals, R., & Dronkers, J. (2016). Tracking and selection on performance. *Applied Economics*, 48(30), 2836–2851.
- Lavy, V. (2008). Do gender stereotypes reduce girls' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, 92(10–11), 2083–2105.
- Lindahl, E. (2007). *Comparing teachers' assessments and national test results—Evidence from Sweden*, IFAU Working Paper no. 2007:24.
- Murnane, J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics*, 77(2), 251–266.
- OECD. (2007). *Thematic review of tertiary education—The Netherlands*, Country Note. Paris: OECD.
- Segal, C. (2008). Classroom behavior. *Journal of Human Resources*, 43(4), 783–814.
- Timmermans, A., Kuyper, H., & van der Werf, G. (2012). *Schooladviezen en onderwijsloopbanen. Voorkomen, risicofactoren en gevolgen van onder- en overadvisering*. Gronings Instituut voor Onderwijs van Onderwijs, Rijksuniversiteit Groningen.
- van Elk, R., van der Steeg, M., & Webbink, D. (2009). *The effect of early tracking on participation in higher education*, CPB Document no. 182. <http://www.cpb.nl/en/publication/effect-early-tracking-participation-higher-education>.
- Woessmann, L. (2004). Institutional comparisons in educational production. *Journal for Institutional Comparisons*, 2(4), 3–6.